

Attorney Docket No.
SONYP031
SCEA03010US00

PATENT APPLICATION

METHOD AND APPARATUS FOR OPTIMIZING CAPTURE DEVICE SETTINGS THROUGH DEPTH INFORMATION

INVENTOR: Richard L. Marks
919 East Hillsdale Boulevard., 2nd Floor
Foster City, CA 94404-2175
Citizen of United States of America

Assignee: Sony Computer Entertainment Inc.

MARTINE & PENILLA, LLP
710 Lakeway Dr., Suite 170
Sunnyvale, California 94085
Telephone (408) 749-6900

METHOD AND APPARATUS FOR OPTIMIZING CAPTURE DEVICE SETTINGS THROUGH DEPTH INFORMATION

by Inventor:

Richard L. Marks

5

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application is related to U.S. Patent Application No. 10/365,120 filed on February 11, 2003 and entitled "METHOD AND APPARATUS FOR REAL TIME MOTION CAPTURE." This application is herein incorporated by reference in its entirety for all purposes.

10

BACKGROUND OF THE INVENTION

1. Field of the Invention

[0002] This invention relates generally to image capture techniques, and more particularly to enhancing a captured image of a scene by adjustments enabled through depth information.

15

2. Description of the Related Art

[0003] Image capture devices, whether cameras or video based devices, typically have a limited contrast ratio, which is a measure of the difference between the darkest and lightest parts of a scene. One exemplary scene may include a person in the shade and the background having bright sunlight. When the background of this scene is exposed correctly for the image capture device, there is little or no detail in the shaded person's face.

20

[0004] Auto-exposure and auto-gain features are commonly used to set brightness levels for the capture device. These features tend to take the entire scene and average it to apply a certain exposure or gain setting. While the averaging may work well for a scene having a great deal of images and colors, this scheme quickly breaks down as the scene has less
5 variety.

[0005] One attempt to address the limited contrast ratio of current capture devices is through the use of a backlight feature. For instance, where there is a bright light source in the background, e.g., sunlight, backlight compensation will take the center of the scene and use that region as the average. By doing this, the center of the scene may be
10 brightened while the bright sunlight on the edges becomes washed out or darkened. The shortcoming with backlight compensation is that the object to be brightened must be in the center of the scene. In addition, a region of the scene is used for computing the average, rather than the actual object itself, which may cause some display artifacts. Furthermore, backlight compensation does not provide a solution where there are
15 multiple foreground images in different regions of the scene. Additionally, with backlight compensation, the foreground object is brightened, however, this is done at the expense of the detail in the background. Thus, a user is required to choose between foreground detail and background detail. Some of these shortcomings may be extended to video capture devices which may be used for interactive entertainment applications. For
20 example, where an image of a user is incorporated into a video game, a bright light source may adversely affect the displayed image as described above. This adverse impact may prevent the tracking of an object of the image in addition to displaying a poor quality image.

[0006] Accordingly, there is a need to solve the problems of the prior art to provide a system and method for producing an image that has the exposure/gain and other related parameters adjusted for both the foreground and background.

SUMMARY OF THE INVENTION

[0007] Broadly speaking, the present invention fills these needs by providing a method and system that enables adjustment of segments of a scene, e.g., foreground and background images, where the foreground and background images are identified through
5 a depth mask. It should be appreciated that the present invention can be implemented in numerous ways, including as a process, a system, or a device. Several inventive embodiments of the present invention are described below.

[0008] In one embodiment, a method for differentiating between foreground objects and background objects within a scene being captured through an image capture device is
10 provided. The method initiates with emitting a ray of light from a light source toward an object of the scene. The method includes opening an aperture cover allowing access to a sensor of the image capture device for reflected light from the light source. Then, the aperture cover is closed after a predefined time, where the predefined amount of time corresponds to a maximum distance traveled by the light. Next, a depth mask identifying
15 objects within a foreground region of the scene is generated based upon the light captured during the predefined time. Then, image capture device parameters are adjusted according to bit values of the depth mask prior to capturing a subsequent corresponding image of the scene.

[0009] In another embodiment, a method for adjusting image capture settings for an
20 image capture device is provided. The method initiates with identifying a scene. Then, an image of the scene is captured. The method includes generating a depth mask of the scene from data defining the image of the scene. Then, pixel values corresponding to objects within any one or both of a foreground region and a background region of the captured image are adjusted based upon bit values of the depth mask.

[0010] In yet another embodiment, an image capture device configured to capture an image of a scene is provided. The image capture device includes depth logic configured to provide a depth mask associated with the scene. The depth mask is configured to distinguish between foreground objects and background objects within the scene. The image capture device also includes image capture logic configured to adjust a characteristic associated with the image based upon a corresponding bit value of the depth mask. The bit value determines whether the respective pixel is associated with one of the foreground objects and the background objects.

[0011] In still yet another embodiment, a system is provided. The system includes a computing device and a display screen in communication with the computing device. The display screen is configured to display an image of a scene. A video capture device in communication with the computing device is included. The video capture device provides scene image data to the computing device for presentation on the display screen. The video capture device includes depth logic and image capture logic. The depth logic is configured to provide a depth mask associated with the scene. The depth mask is configured to distinguish between foreground objects and background objects within the scene. The image capture logic is configured to adjust a characteristic associated with each pixel of the image data based upon corresponding data of the depth mask. The data of the depth mask determines a relative distance between an object associated with the corresponding pixel and the video capture device.

[0012] Other aspects and advantages of the invention will become apparent from the following detailed description, taken in conjunction with the accompanying drawings, illustrating by way of example the principles of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

[0013] The invention, together with further advantages thereof, may best be understood by reference to the following description taken in conjunction with the accompanying drawings.

5 [0014] Figure 1 is a simplified schematic diagram illustrating a scene having foreground and background objects, which is captured through an image capture device and subsequently displayed in accordance with one embodiment of the invention.

[0015] Figure 2 is a simplified schematic diagram illustrating the generation of a depth mask for use in discerning between foreground and background objects in accordance
10 with one embodiment of the invention.

[0016] Figures 3A and 3B are simplified schematic diagrams illustrating the amount of detail enabled in defining foreground objects in accordance with one embodiment of the invention.

[0017] Figure 4 is a simplified schematic diagram illustrating a captured image which is
15 enhanced through a generated mask to define background and foreground images in accordance with one embodiment of the invention.

[0018] Figure 5 is a simplified schematic diagram of an interactive entertainment system which utilizes the mask generation in order to more effectively track a user in accordance with one embodiment of the invention.

20 [0019] Figure 6 is a simplified schematic diagram of an image capture device in accordance with one embodiment of the invention.

[0020] Figure 7 is an alternative schematic diagram of an image capture device having logic configured to differentiate between foreground and background images in the invention.

[0021] Figure 8 is a flow chart diagram illustrating the method operations for adjusting image capture settings for an image capture device in accordance with one embodiment of the invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

[0022] An invention is disclosed for a system and method for differentiating between foreground and background objects of a scene and subsequently adjusting image or video characteristics based upon whether the objects are located in the foreground or background. Alternatively, the image or video characteristics may be adjusted based upon the relative distance between the objects and the image capture device. In the following description, numerous specific details are set forth in order to provide a thorough understanding of the present invention. It will be apparent, however, to one skilled in the art that the present invention may be practiced without some or all of these specific details. In other instances, well known process steps have not been described in detail in order not to unnecessarily obscure the present invention.

[0023] The embodiments of the present invention provide a method and system that eliminates the user from having to choose between foreground and background objects of a scene. Through the use of depth information, the scene may be segmented into regions of different depths. In addition, the depth information allows for the definition of an exact outline of the image, thereby providing a specific and accurate mechanism for controlling image capture device parameters, e.g., exposure/gain/brightness/gain and focus. The segmentation based upon depth information makes it possible to assign different parameter values to different pixel regions for a digital capture device. Thus, an image having the exposure/gain adjusted properly for both the foreground and background is enabled through the embodiments described below. The segmentation based upon depth is captured through a foreground/background depth mask which may be created through an image capture device having depth capability or through a light pulse/flash with a time of flight cut-off technique, both discussed in more detail below.

[0024] Figure 1 is a simplified schematic diagram illustrating a scene having foreground and background objects which is captured through an image capture device and subsequently displayed in accordance with one embodiment of the invention. Image capture device 100 is configured to capture an image of a scene in which a person 102 is in the foreground and background scenery 104. The captured image of the scene is then displayed on display panel 106. Display panel 106 may be a display panel affixed to image capture device 100, e.g., a liquid crystal display (LCD) panel where the image capture device is a digital camera or camcorder. Alternatively, display panel 106 may be remote from image captures device 100, e.g., a television screen where the image capture device is a webcam used in conjunction with a computing device, such as a game console. As will be described in more detail below, foreground image 102 and background scenery 104 are capable of having their corresponding image or video characteristics independently compensated irrespective of their position in either the foreground or the background. While a single foreground image 102 is shown in Figure 1, it should be appreciated that multiple foreground images may be captured. The image or video characteristics for each of the multiple foreground images may be independently adjusted based upon depth information. As used herein, image or video characteristics may refer to brightness, exposure, gain, focus and other suitable characteristics capable of being adjusted for image display. It should be appreciated that image or video characteristics may be referred to simply as characteristics and correspond to the inherent image data which improves the display quality of the image data through the embodiments described herein. Additionally, image capture device 100 may be a digital still camera, a single lens reflex camera, a video capture device, such as a web cam or camcorder, or any other suitable image capture device.

[0025] Image capture device 100, of Figure 1, is capable of generating and utilizing a mask in order to identify objects as being within a foreground or background region as will be described in more detail below. This mask can then be used in order to compensate for the foreground and background regions, in order to provide a subsequent display which shows details for objects within both regions. For example, the use of backlight compensation to reduce the impact of a bright light source on the scene, such as sun 108, causes details to be defined for foreground objects, i.e., object 102, while the background images are washed out. While sun 108 is illustrated on display panel 106, it is shown for exemplary purposes and a scene need not include an actual image of the sun to be adversely impacted by the light originating from the sun. Without backlight compensation, foreground objects would be darkened and lose their corresponding detail in the resulting display. With the depth mask capability described in more detail below, the exact location of foreground and background and background objects in the scene may be determined. This location may be translated to a resulting image of the scene in order to manipulate corresponding pixel values to enhance the resulting image. In addition, image capture device settings, which include mechanical and electrical settings that affect the image or video characteristics of the resulting image, may be adjusted in order to provide optimized settings for the scene.

[0026] Where image capture device 100 is a video capture device, e.g., a web cam, the enhanced functionality enabled through the depth information provided by the mask may be applied to frames of the captured video in order to improve or enhance the image display. For example, where the video capture device is used to track an object or person subsequently incorporated into an interactive entertainment application, the mask may be applied to prevent difficulty encountered when tracking the object or person in the

presence of a bright light source. With respect to a video game application in the home environment, such as the EYETOY™ application owned by the assignee, a user being tracked and incorporated into the video game may be positioned in front of a window. As explained below in more detail with reference to Figure 4, if the window is allowing light from a bright light source through the window, then the user may become washed out and the window will become the focus of the capture device. It should be appreciated that backlight compensation techniques will not be effective here if the user is not in the center of the capture region.

[0027] Figure 2 is a simplified schematic diagram illustrating the generation of a depth mask for use in discerning between foreground and background objects in accordance with one embodiment of the invention. It should be noted that the terms “depth mask” and “mask” are interchangeable as used herein and may include multiple depth layers. For example, the foreground and the background represent 2 depth layers, however, the scene may be segmented into more than two depth layers. Image capture device 100 includes light source 110. In one embodiment, light source 110 sends out a burst or pulse of light which is reflected by foreground objects 114 and 116. This reflected light is eventually captured by a sensor located behind lens 112 of image capture device 100. Of course, light source 110 may be a flash commonly used for cameras. One skilled in the art will appreciate that the sensor may be located anywhere on image capture device 100 that is capable of receiving the reflected light from the foreground objects within the scene for a defined time period.

[0028] As the speed of light is known, image capture device 100 of Figure 2 may be configured to pulse the burst of light from light source 110 and open an aperture of image capture device 100 so that the reflected light from the foreground objects is received. The

aperture will stay open for a predefined amount of time. The predefined amount of time is set so that light traveling from light source 110 and reflected back to image capture device 100, travels a defined maximum amount of distance. The maximum distance from image capture device 100 is illustrated as line 117. Therefore, any light which is reflected
5 from a source past line 117 will not be received by image capture device as the aperture is closed prior to this reflected light reaching the sensor of the image capture device. Of course, the ambient light, i.e., the light not generated by the burst of light from the light source, is subtracted from the received light.

[0029] Various techniques may be used for the determining the foreground objects
10 through the time of flight. One technique is through the use of a frequency of light not present in the ambient light. Alternatively, an image of the scene may be taken without the light on, then an image taken with the light from the light source. The light generated by the light source may then be determined by subtracting away the light not generated from the light source, i.e., the image taken without the light on, from the image taken
15 with the light source. In yet another alternative, the amount of light reflected from the light source may be distinguished from ambient light by establishing a threshold of how much light must strike each pixel. Thus, a value which is less than the threshold would not be considered as light originating from the device and values greater than or equal to the threshold would be considered as originating from the light source of the device. Still
20 yet another alternative that employs the use of a modulated light source. Here, the light from the light source is generated in a modulated format, e.g., a sine wave. The frequency chosen depends upon a range where no more than one period of the modulation covers the entire range from the light source and back to the device.

[0030] In one embodiment, the maximum amount of distance is defined as about four meters from the image capture device. From this data, a depth mask is created and stored in memory of the image capture device. This depth mask may then be used in conjunction with a simultaneous or subsequent captured image of the scene in order to
5 compensate for the image or video characteristics for the foreground and background objects accordingly. It will be apparent to one skilled in the art that light source 110 may emit any suitable wavelength of light. In one embodiment, infrared light is emitted from light source 110.

[0031] In another embodiment, the depth mask defined through the reflected light is a
10 binary bit mask. Here, a first logical value is assigned to locations in the mask associated with foreground images, while a second logical value is assigned to locations associated with background images. Thus, where image capture device 100 is a digital device, pixel data for an image associated with the depth mask may be manipulated to adjust the brightness of the foreground and background images. Where image capture device is a
15 traditional camera, foreground and background images may be detected through the burst of light scheme described above. Based on the detection of the foreground and background images, the exposure, gain, brightness, focus, etc., settings of the camera may be adjusted prior to taking a picture of the scene. As mentioned above, the aperture size may be changed to manipulate the amount of light received by the image capture device.
20 Of course, other mechanical and electrical settings may be adjusted where the mechanical or electrical settings impact the resulting photograph quality. Thus, both the foreground and background properties may be adjusted rather than having to choose between the foreground and the background.

[0032] Figures 3A and 3B are simplified schematic diagrams illustrating the amount of detail enabled in defining foreground objects in accordance with one embodiment of the invention. Figure 3A illustrates display screen 120 having a foreground object defined through rectangular region 122. Figure 3B shows display screen 120 illustrating a foreground object 124 in which a mask has been defined, as described herein, in order to capture the exact outline of the foreground image. That is, with current auto focus, auto gain, backlight compensation techniques, the center of a scene in which an image capture device is targeting, is generally represented as an area and is incapable of outlining the exact image. Thus, as illustrated in Figure 3A, rectangular region 122 includes the foreground object as well as other image data. Furthermore, the foreground object must be within a center region of the image or the auto focus, auto gain, or backlight compensation features will not work. In contrast, the depth mask captures any foreground object irrespective of its location within the scene. Moreover, the foreground object is captured without any additional image data being included. As mentioned above, for a digital device, the image or video characteristics for any foreground object may be manipulated by adjusting pixel values. With respect to a traditional film camera, the gain, exposure, focus, and brightness may be manipulated through mechanical or electrical adjustments responsive to the depth mask.

[0033] Figure 4 is a simplified schematic diagram illustrating a captured image which is enhanced through a mask generated to define background and foreground images in accordance with one embodiment of the invention. Here, image scene 128 may be a scene captured through an image capture device such as a video cam or a web cam for an interactive gaming application where participant 130 is incorporated into the interactive gaming application. An exemplary interactive gaming application is the EYETOY™

interactive game application. Here, participant 130 is standing in front of a web cam or some other suitable video capture device. Behind participant 130 is window 132. It should be appreciated that where bright light is shining through window 132, the resulting image of participant 130 captured by the image capture device will become darkened. In

5 an interactive video game application where tracking the user is important, the tracking will become difficult where the bright light darkens the image of the user. Thus, where the video cam incorporates the embodiments described herein, the user will be able to be tracked more easily. That is, a mask generated as described above, may be used to manipulate the pixel values to reduce the brightness.

10 **[0034]** Figure 5 is a simplified schematic diagram of an interactive entertainment system which utilizes the generated mask in order to more effectively track a user in accordance with one embodiment of the invention. Here, image capture device 100 is configured to capture an image of user 134 in order for the user's image to be displayed on display screen 136. Image capture device 100 is in communication with computing device 138,

15 which in turn, is in communication with display screen 136. As can be seen, image 135 of user 134 is displayed on display screen 136. Thus, as user 134 moves, this movement is captured through image capture device 100 and displayed on display screen 136 in order to interact with the entertainment application. As mentioned above, the image capture device is configured to compensate for bright light entering through window 132.

20 **[0035]** Still referring to Figure 5, image capture device 100 is a video capture device. Here, the pixel data associated with each video frame may be adjusted according to a corresponding depth mask. In one embodiment, a depth mask is generated for each video frame. In another embodiment, the depth mask is generated every x number of frames, where x may be any integer. For the frames not associated with a mask in this

embodiment, the image or video characteristics from the last previous frame associated with a mask are applied to the frames not associated with a mask. Thus, the image or video characteristics may be frozen for a certain number of frames until a new mask is generated. It will be apparent to one skilled in the art that the processing for the

5 functionality described herein may be performed by a processor of computing device 138. However, the depth mask may be generated by image capture device 100 and stored in memory of the image capture device. Of course, the image capture device would contain a microprocessor for executing the functionality for generating the depth mask and adjusting the image or video characteristics or adjusting the device parameters.

10 **[0036]** Image capture device 100 of Figure 5 may generate the mask through the techniques described with reference to Figure 2, however, image capture device 100 may alternatively include depth capturing logic, such as 3DV SYSTEM's ZCAMTM or similar products commercially available through CANESTATM. The depth capturing logic includes an image sensor that captures the depth value of each pixel in a scene in order to

15 create a depth mask to be used as discussed herein. It should be noted that while a single user 134 is depicted in Figure 5, it should be noted that multiple users may be incorporated in the embodiments described here. Since the depth mask enables adjustment of both foreground and background object image or video characteristics, it is not required that user 134 be located in the middle or any other particular area of the

20 capture region for image captures device 100. It should be further appreciated that one exemplary system represented by Figure 5 is the EYETOYTM system mentioned above.

[0037] Figure 6 is a simplified schematic diagram of an image capture device in accordance with one embodiment of the invention. Image capture device 100 includes depth logic 140, image capture device logic 142, and memory 144 all in communication

with each other. As described herein, depth logic 140 includes circuitry configured to generate a mask in order for image capture device 100 to enhance a captured image with the assistance of the depth information. For example, depth logic 140 may generate the mask in order to differentiate between foreground and background objects within an image scene, and this mask will be stored in memory 144. Then, a corresponding scene of the image that is captured and processed by image capture device logic 142 will be enhanced. That is, certain image or video characteristics are manipulated as described herein depending on whether an object within the scene is located in the foreground or background, as determined by the depth mask. In one embodiment, depth logic 140 is activated by button 141 or some other suitable activation mechanism. Thus, a user has the option of activating the depth logic for enhanced image presentation, or bypassing the image presentation.

[0038] Figure 7 is an alternative schematic diagram of an image capture device having logic configured to differentiate between foreground and background images in the invention. Image capture device 100 includes lens 150 behind which is charged coupled device (CCD) 152. Depth logic 140, microprocessor unit (MPU) 148, and memory 144 are also included. Image capture device 100 includes display panel 154. It will be apparent to one skilled in the art that while image capture device 100 is depicted as a digital camera in Figure 7, the invention is not limited to a digital camera. Depth logic module 140 may be included in a video capture device in order to adjust image or video characteristics of each frame or every x^{th} frame.

[0039] Figure 8 is a flow chart diagram illustrating the method operations for adjusting image capture settings for an image capture device in accordance with one embodiment of the invention. The method initiates with operation 160 where a scene is identified.

Here, an image capture device may be used to identify a scene defined by a capture region. Of course, the image capture device may be a video capture device. The method then advances to operation 162 where a depth mask of the scene is generated for segmentation of foreground and background regions. In one embodiment, the depth mask is generated by pulsing light and capturing reflections from an object within a certain distance as described with reference to Figure 2. Here the light may be infrared light. In another embodiment, the image capture device includes depth logic capable of capturing a depth value for each pixel. One exemplary image capture device with depth logic is the ZCAMTM mentioned above. The method then proceeds to operation 164 where an image of the scene is captured and this captured image corresponds to the depth mask. It should be appreciated that for the ZCAMTM embodiment, operations 162 and 164 are performed simultaneously. The method then moves to operation 166 where pixel values of objects within either, or both, of the foreground and background regions of the captured image are adjusted. This adjustment is based upon the depth mask defined above.

[0040] For example, the depth mask may be defined through bit values where a first bit value is assigned to foreground objects and a second bit value is assigned to background objects. The adjustment then enhances the brightness of foreground objects while decreasing the brightness of background objects where a bright light source exists in one embodiment. Where the image capture device is not a digital device, e.g., a SLR camera, mechanical or electrical adjustments of the image capture device parameters may be made as a result of the foreground and background objects identified by the bit mask. These mechanical or electrical adjustments may include defining an aperture size corresponding to a certain exposure level, lens settings for a particular focus level, etc. In another embodiment, the pixel values are adjusted according to depth information included with

the image data, i.e., distance information tagged to each pixel of the image data. One skilled in the art will appreciate that the aperture size may be controlled mechanically or electronically. The electronic control may be performed through a sensor on a chip. Thus, each pixel adjusted separately with the electronic control.

5 **[0041]** In summary, an image capture device capable of generating a depth mask for corresponding segments of a scene is provided. It should be appreciated that while the invention has been described in terms of the background and foreground segments (2 layers) of a scene, the embodiments described herein may be extended to any number of layers of the scene. Through the depth mask, the image or video characteristics for an
10 image may be selectively adjusted irrespective of where an object is located in the scene. Furthermore, the capture device described herein enables enhanced functionality for interactive entertainment applications. For example, with respect to a video game application, where a user is tracked to incorporate his image into the video game, the capture device described above enables for enhanced tracking of the user. The user is
15 free to move anywhere in the capture region and is not limited to one area, such as a center region. Additionally, as the user moves in front of a bright light source, e.g., sunlight coming through a window, the detail of the user's image is not lost. With respect to a video capture device, the adjustments may be applied every interval of frames in order to avoid constant adjustments from occurring. For example, if a user briefly holds
20 up a black piece of paper in front of him, the frame interval delay will prevent the user from suddenly turning darker. Also, if the user temporarily leaves the field of view of the image capture device and comes back, the adjustment and re-adjustment of the scene is avoided.

[0042] It should be appreciated that the embodiments described above may be extended to other systems in addition to an interactive entertainment input device, i.e., the EYETOY™ system capture device. For example, the video capture device may be used in a videoconferencing system to provide enhanced video images for the conference.

5 Here, the capture device may not be used for tracking purposes, but for the enhancement of the image or video characteristics enabled through the depth information.

[0043] The invention may employ various computer-implemented operations involving data stored in computer systems. These operations are those requiring physical manipulation of physical quantities. Usually, though not necessarily, these quantities take
10 the form of electrical or magnetic signals capable of being stored, transferred, combined, compared, and otherwise manipulated. Further, the manipulations performed are often referred to in terms, such as producing, identifying, determining, or comparing.

[0044] Any of the operations described herein that form part of the invention are useful machine operations. The invention also relates to a device or an apparatus for performing
15 these operations. The apparatus may be specially constructed for the required purposes, or it may be a general purpose computer selectively activated or configured by a computer program stored in the computer. In particular, various general purpose machines may be used with computer programs written in accordance with the teachings herein, or it may be more convenient to construct a more specialized apparatus to perform the required
20 operations.

[0045] Although the foregoing invention has been described in some detail for purposes of clarity of understanding, it will be apparent that certain changes and modifications may be practiced. Accordingly, the present embodiments are to be considered as illustrative

and not restrictive, and the invention is not to be limited to the details given herein, but may be modified within the scope and equivalents of the description.

What is claimed is: